Europäisches Patentamt
European Patent Office
Office européen des brevets

EP 99 4416

5

Bescheinigung Certificate Attestation

Die angehefteten Unterlagen stimmen mit der ursprünglich eingereichten Fassung der auf dem nächsten Blatt bezeichneten europäischen Patentanmeldung überein.

The attached documents are exact copies of the European patent application described on the following page, as originally filed.

Les documents fixés à cette attestation sont conformes à la version initialement déposée de la demande de brevet européen spécifiée à la page suivante.

Patentanmeldung Nr. Patent application No. Demande de brevet n°

98305033.7

PRIORITY DOCUMENT

SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

Der Präsident des Europäischen Patentamts;
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets
p.o.

H.-P. Dietenhofer

DEN HAAG, DEN
THE HAGUE,
LA HAYE, LE 07/06/99

Europäisches Patentamt
European Patent Office
Office européen des brevets

Blatt 2 der Bescheinigung
Sheet 2 of the certificate
Page 2 de l'attestation

Anmeldung Nr.:
Application no.: 98305033.7
Demande n°:

Anmeldetag:
Date of filing: 25/06/98
Date de dépôt:

Anmelder:
Applicant(s):
Demandeur(s):
GIST-BROCADES B.V.
2600 MA Delft
NETHERLANDS

Bezeichnung der Erfindung:
Title of the invention:
Titre de l'invention:
Proprionibacterium vector

In Anspruch genommene Priorität(en) / Priority(ies) claimed / Priorité(s) revendiquée(s)

Staat: State: Pays:	Tag: Date: Date:	Aktenzeichen: File no. Numéro de dépôt:

Internationale Patentklassifikation:
International Patent classification:
Classification internationale des brevets:
C12N15/31, C12N15/63

Am Anmeldetag benannte Vertragstaaten:
Contracting states designated at date of filing: AT/BE/CH/CY/DE/DK/ES/FI/FR/GB/GR/IE/IT/LI/LU/MC/NL/PT/SE
Etats contractants désignés lors du depôt:

Bemerkungen:
Remarks:
Remarques:

PROPIONIBACTERIUM VECTOR

This invention relates to an endogenous plasmid of Propionibacterium, vectors derived from it and the use of these vectors to express (heterologous) proteins in bacteria, especially Propionibacteria. In particular transformed bacteria can be used to produce, by fermentation, vitamin B12.

Introduction

Propionibacteria are Gram-positive bacteria capable of producing valuable compounds in a variety of industrial processes. For example, several Propionibacterium species are known to produce vitamin B12 (Cobalamin) in large scale fermentations processes. Other species are used in dairy applications such as cheese manufacturing where they contribute, and in many cases even are mainly responsible, for the specific flavour and texture of the cheese. Many Propionibacterium species are considered safe for inclusion, as live organisms, into food and animal feed.

To be able to fully exploit the biotechnological potential of Propionibacterium, efficient and flexible genetic engineering techniques are required. Such techniques rely on the availability of a suitable plasmid to express a protein from a heterologous gene in Propionibacterium.

EP-A-0400931 refers to an endogenous plasmid (pTY-1) from Propionibacterium pentosaceum (ATCC 4875) but does not describe its sequence or exemplify how it may be used to express a heterologous gene.

JP 8-56673 refers to the plasmid pTY-1 for producing vitamin B12 but does not provide any evidence that the plasmid remains as a freely replicating extrachromosomal element nor that the plasmid is stable inside the transformed cells.

The invention therefore seeks to provide vectors that are more efficient than those in the prior art, and

can remain extrachromosomal and/or are stable. In particular the invention aims to provide an efficient vector for the cloning or expression of Propionibacterium or foreign genomic fragments or genes into a (Propionibacterium) host strain. This may enable host specific restriction enzymes to be circumvented and thereby avoid the host treating the plasmid as a foreign polynucleotide.

Summary of the invention

Accordingly, the present invention in a first aspect provides a polynucleotide comprising a sequence capable of hybridising selectively to

(i) SEQ ID NO: 1 or the complement thereof;

(ii) a sequence from the 3.6 kb plasmid of Propionibacterium freudenreichii CBS 101022; or

(iii) a sequence from the 3.6 kb plasmid of Propionibacterium freudenreichii CBS 101023.

The polynucleotide may encode (at least part of) the amino acid sequence of SEQ ID NO: 2 or SEQ ID No: 3 (which forms the second aspect). SEQ ID NO: 1 sets out the DNA sequence of the endogenous plasmid of Propionibacterium LMG 16545 which the inventors have discovered. The first coding sequence runs from nucleotide 273 to nucleotide 1184. The predicted amino acid sequence of this coding sequence is shown in SEQ ID NO: 2. The second coding sequence runs from nucloetides 1181 to 1483. The predicted amino acid sequence of this coding sequence is shown in SEQ ID No: 3.

The inventors have screened a large collection of Propionibacterium isolates and identified two strains harboring cryptic plasmids with a size of 3.6 kb. One of the strains is Propionibacterium freudenreichii LMG 16545 which was deposited at Centraalbureau voor Schimmelcultures (CBS), Oosterstraat 1, Postbus 273, NL-3740 AG Baarn, Netherlands, in the name of Gist-brocades B.V. of Wateringseweg 1, P.O. Box 1, 2600 MA Delft, The

Netherlands, on 20 June 1998 under the terms of the Budapest Treaty and was given accession number CBS 101022. The other strain is Propionibacterium freudenreichii LMG 16546 which was deposited by the same depositor on 20 June 1998 under the terms of the Budapest Treaty also at Centraalbureau voor Schimmelcultures and was given accession number CBS 101023.

Through full characterization and computer assisted analysis of the nucleotide sequence of LMG 16545 the inventors have been able to identify insertion sites for foreign DNA fragments. This allows plasmids to be constructed using the sequence information which are still capable of autonomous replication in Propionibacterium.

Surprisingly, the inventors found that an erythromycin resistance gene from the actinomycete *Saccharopolyspora erythraea* is efficiently expressed in Propionibacterium and thus can be used as a selection marker for transformed cells.

The inventors also constructed bifunctional vectors, stably maintainable and selectable in both *E.coli* and Propionibacterium. This can allow the use of *E. coli* for vector construction, as well as functional expression of homologous or heterologous genes in Propionibacterium. Vector construction using *E. coli* is comparatively easy and can be done quickly.

The polynucleotide of the invention may be autonomously replicating or extrachromosomal , for example in a bacterium such as a Propionibacterium.

Thus in a second aspect the invention provides a vector which comprises a polynucleotide of the invention.

The invention also provides a process for the preparation of a polypeptide, the process comprising cultivating a host cell transformed or transfected with a vector of the invention under conditions to provide for expression of the polypeptide.

In another aspect the invention provides a

polypeptide which comprises the sequence set out in SEQ ID NO: 2 or 3 or a sequence substantially homologous thereto, or a fragment of either sequence. Polynucleotides also include those encoded by a polynucleotide of the first aspect.

Detailed Description of the Invention

A polynucleotide of the invention may be capable of hybridising selectively with the sequence of SEQ ID NO: 1, or a portion of SEQ ID No: 1, or to the sequence complementary to that sequence or portion of the sequence. The polynucleotide of the invention may be capable of hybridising selectively to the sequence of the 3.6 kb plasmid of P. freudenreichii CBS 101022 or CBS 101023, or to a portion of the sequence of either plasmid. Typically, a polynucleotide of the invention is a contiguous sequence of nucleotides which is capable of selectively hybridizing to the sequence of SEQ ID. No: 1 or of either 3.6 kb plasmid, or portion of any of these sequences, or to the complement of these sequences or portion of any of these sequences.

A polynucleotide of the invention and the sequence of SEQ ID NO: 1 or either of the 3.6 kb plasmids, or portion of these sequences, can hydridize at a level significantly above background. Background hybridization may occur, for example, because of other polynucleotides present in the preparation. The signal level generated by the interaction between a polynucleotide of the invention and the sequence of SEQ ID NO: 1 or of either 3.6 kb plasmid, or portion of these sequences, is typically at least 10 fold, preferably at least 100 fold, as intense as interactions between other polynucleotides and the coding sequence of SEQ ID NO: 1 or of either 3.6 kb plasmid, or portion of these sequences. The intensity of interaction may be measured, for example, by radiolabelling the probe, e.g. with ^{32}P. Selective hybridisation is typically achieved using conditions of

medium to high stringency (for example 0.03M sodium chloride and 0.03M sodium citrate at from about 50°C to about 60°C).

Polynucleotides included in the invention can be generally at least 70%, preferably at least 80 or 90% and more preferably at least 95%, homologous to the sequence of SEQ ID No: 1 or its complement or of either 3.6 kb plasmid over a region of at least 20, preferably at least 30, for instance at least 40, 60 or 100 or more contiguous nucleotides.

Any combination of the above mentioned degrees of homology and minimum sizes may be used to define polynucleotides of the invention, with the more stringent combinations (i.e. higher homology over longer lengths) being preferred. Thus for example a polynucleotide which is at least 80% homologous over 25, preferably over 30 nucleotides forms one embodiment of the invention, as does a polynucleotide which is at least 90% homologous over 40 nucleotides.

The portions referred to above may be the coding sequences of SEQ ID No: 1 or of either 3.6 kb plasmid. Other preferred portions of SEQ ID No: 1 are the replication origin, promoter or regulatory sequences, or sequences capable of effecting or assisting autonomous replication in a host cell, such as a Propionibacterium.

Polynucleotides of the invention may comprise DNA or RNA. They may also be polynucleotides which include within them synthetic or modified nucleotides. A number of different types of modification to polynucleotides are known in the art. These include methylphosphonate and phosphorothioate backbones, addition of acridine or polylysine chains at the 3' and/or 5' ends of the molecule. For the purposes of the present invention, it is to be understood that the polynucleotides described herein may be modified by any method available in the art. Such modifications may be carried out in order to enhance the *in vivo* activity or lifespan of

polynucleotides of the invention.

Polynucleotides of the invention may be used to produce a primer, e.g. a PCR (polymerase chain reaction) primer, a primer for an alternative amplification reaction, a probe e.g. labelled with a revealing label by conventional means using radioactive or non-radioactive labels, or the polynucleotides may be incorporated into vectors. Such primers, probes and other fragments will be at least 15, preferably at least 20, for example at least 25, 30 or 40 nucleotides in length.

Polynucleotides such as a DNA polynucleotide and primers according to the invention may be produced recombinantly, synthetically, or by any means available to those of skill in the art. They may also be cloned by standard techniques. The polynucleotides are typically provided in isolated and/or purified form.

In general, primers will be produced by synthetic means, involving a step wise manufacture of the desired nucleic acid sequence one nucleotide at a time. Techniques for accomplishing this using automated techniques are readily available in the art.

Longer polynucleotides will generally be produced using recombinant means, for example using PCR cloning techniques. This will involve making a pair of primers (e.g. of about 15-30 nucleotides) to the region of SEQ ID No: 1 or of either 3.6 kb plasmid which it is desired to clone, bringing the primers into contact with DNA obtained from a Propionibacterium, performing a polymerase chain reaction under conditions which bring about amplification of the desired region, isolating the amplified fragment (e.g. by purifying the reaction mixture on an agarose gel) and recovering the amplified DNA. The primers may be designed to contain suitable restriction enzyme recognition sites so that the amplified DNA can be cloned into a suitable cloning vector. Such techniques may be used to obtain all or part of SEQ ID No: 1 or either 3.6 kb plasmid.

Although in general the techniques mentioned herein are well known in the art, reference may be made in particular to Sambrook *et al*, 1989.

Polynucleotides which are not 100% homologous to SEQ ID No: 1 or either 3.6 kb plasmid but fall within the scope of the invention can be obtained in a number of ways.

Homologous polynucleotides of SEQ ID NO: 1 or of either 3.6 kb plasmid may be obtained for example by probing genomic DNA libraries made from a range of Propionibacteria, such as P.freudenreichii, P.jensenii, P.acidipropionici, or other strains of bacteria of the class Actinomycetes, or other gram positive bacteria, or those that are G: C rich. All these organisms are suitable sources of homologous or heterologous genes, promoters, enhancers, or host cells, for use in the invention.

Such homologues and fragments thereof in general will be capable of selectively hybridizing to the coding sequence of SEQ ID NO: 1 or its complement or of either 3.6 kb plasmid. Such sequences may be obtained by probing genomic DNA libraries of the Propionibacterium with probes comprising all or part of the coding sequence SEQ ID NO: 1 or of either 3.6 kb plasmid under conditions of medium to high stringency (for example 0.03M sodium chloride and 0.03M sodium citrate at from about 50°C to about 60°C).

Homologues may also be obtained using degenerate PCR which will use primers designed to target conserved sequences within the homologues. Conserved sequences can be predicted from aligning SEQ ID No: 1 or the sequence of either 3.6 kb plasmid with their homologues. The primers will contain one or more degenerate positions and will be used at stringency conditions lower than those used for cloning sequences with single sequence primers against known sequences.

Alternatively, such polynucleotides may be obtained

by site directed mutagenesis of SEQ ID No: 1 or of either 3.6 kb plasmid, or their homologues. This may be useful where for example silent codon changes are required to sequences to optimise codon preferences for a particular host cell in which the polynucleotide sequences are being expressed. Other sequence changes may be desired in order to introduce restriction enzyme recognition sites, or to alter the property or function of the polypeptides encoded by the polynucleotides.

The invention includes double stranded polynucleotides comprising a polynucleotide sequence of the invention and its complement.

Polynucleotides or primers of the invention may carry a revealing label. Suitable labels include radioisotopes such as ^{32}P or ^{35}S, enzyme labels, or other protein labels such as biotin. Such labels may be added to polynucleotides or primers of the invention and may be detected using techniques known *per se*.

Polynucleotides or primers of the invention or fragments thereof labelled or unlabelled may be used by a person skilled in the art in nucleic acid-based tests for detecting or sequencing a polynucleotide of the invention, in a sample.

Polynucleotides of the invention include variants of the sequence of SEQ ID NO: 1 or of either 3.6 kb plasmid which are capable of autonomously replicating or remaining extrachromosomally in a host cell. Such variants may be stable in a bacterium such as a Propionibacterium.

Generally the polynucleotide will comprise the replication origin and/or coding region(s) of SEQ ID No: 1 or of either 3.6 kb plasmid, or homologues of these sequences discussed above. A polynucleotide of the invention which is stable in a host cell, such as Propionibacterium, or *E. coli* is one which is not lost from the host within five generations, such as fifteen generations, preferably thirty generations. Generally

such a polynucleotide would be inherited by both daughter cells every generation.

The polynucleotide may comprise a promoter or an origin of replication (e.g. upstream of any sequences encoding for a replication protein).

The polynucleotide of the invention can be transformed or transfected into a bacterium, such as a Propionibacterium, or *E. coli* by any suitable method. Such methods are disclosed in Sambrook *et al*, 1989.

The polynucleotide of the invention may be present in a bacterium at a copy number of 5 to 500, such as 10 to 100.

The polynucleotide of the invention may be capable of autonomous replication in a bacterium other than a Propionibacterium. Such a bacterium may be *E. coli*, or a gram positive or G:C rich bacteria or one of the class Actinomycetes. Such a polynucleotide will generally comprise sequences which enable the polynucleotide to be autonomously replicated in that bacterium. Such sequences can be derived from plasmids which are able to replicate in that bacterium. A polynucleotide of the invention may be one which has been produced by replication in a Propionibacterium. Alternatively the polynucleotide of the invention may have been produced by replication in another bacterium, such as *E. coli*. The polynucleotide may be able to circumvent the host restriction systems of Propionibacterium.

A second aspect of the invention relates to a vector comprising a polynucleotide of the first aspect. The vector may be capable of replication in a host cell, such as a bacterium, for examples Actinomycetes, e.g. Propionibacterium or *E. coli*. The vector may be a linear polynucleotide or, more usually, a circular polynucleotide. The vector may be a hybrid of the polynucleotide of the invention and another vector. The other vector may be an E. coli vector, such as pBR322, pUC, R1, ColD or rSF1010.

The polynucleotide or vector of the invention may be a plasmid. Such a plasmid may have a restriction map the same as or substantially similar to the restriction maps shown in Figure 1,2a or 2b.

The polynucleotide or vector may have a size 1 of kb to 20 kb, such as from 2 to 10 kb, optimally from 3 to 7 kb.

The polynucleotide or vector may comprise multiple functional cloning sites. Such cloning sites generally comprise the recognition sequences of restriction enzymes. The polynucleotide or vector may comprise the sequence shown in SEQ ID No: 1 which contains restriction for enzyme recognition sites for *Eco*RI, *Sac*I, *AlwN*1, *Bsm*I, *Bsa BI*, *Bcl*I, *Apa*I, *Hind*III, *Sal*I, *Hpa*I, *Pst*I, *Sph*I, *Bam*HI, *Acc*65I, *Eco*RV and *Bgl*III. The polynucleotide or vector may thus comprise one, more than one or all of these restriction enzyme sites, suitably in the order shown in the Figures.

Preferably, when present in a bacterium, such as a Propionibacterium or *E. coli*, the polynucleotide or vector of the invention does not integrate into the chromosome of the bacterium. Generally the polynucleotide or vector does not integrate within 5 generations, preferably 20 or 30 generations.

The polynucleotide or vector may be an autonomously replicating plasmid that can remain extrachromosomal inside a host cell, which plasmid is derived from an endogenous Propionibacterium plasmid, and when comprising a heterologous gene (to the vector) is capable of expressing that gene inside the host cell. The term "derived from" means that the autonomously replicating plasmid includes sequence the same as the polynucleotide of the invention.

The vector of the invention may comprise a selectable marker. The selectable marker may be one which confers antibiotic resistance, such as ampicillin, kanamycin or tetracylin resistance genes. The selectable

marker may be an erythromycin resistance gene. The erythromycin resistance gene may be from an Actinomycete, such as Saccharopolyspora erythraea, for example from Saccharopolyspora erythraea NRRL2338. Other selectable markers which may be present in the vector include chloramphenicol, thiostrepton, viomycin, neomycin, apramycin, hygromycin,bleomycin or streptomycin.

The vector of the invention may be an expression vector. Such an expression vector may comprise a heterologous gene (which does not naturally occur in the host cell, e.g. Propionibacteria), or an endogenous or homologous gene of the host cell, e.g. Propionibacteria. In the expression vector the gene to be expressed is operably linked to a control sequence which is capable of providing for the expression of the gene in a host cell.

The term "operably linked" refers to a juxtaposition wherein the components described are in a relationship permitting them to function in their intended manner. A controlled sequence "operably linked" to a coding sequence is ligated in such a way that expression of the coding sequence is achieved under conditions compatible with the control sequences.

The heterologous or endogenous gene may be inserted between nucleotides 1 and 200 or between nucleotides 1500 to 3555 of SEQ ID No: 1 or at an equivalent position in a homologous polynucleotide.

Examples of such genes include homologous or endogenous genes such as for elongation factors, promoters and replication proteins. Other genes (which may be heterologous to the host) include those encoding for or assisting in the production of nutritional factors, immunomodulators, hormones, proteins and enzymes (e.g. proteases, peptidases, lipases), texturing agents, flavouring substances (e.g. diacetyl, acetoin), gene clusters, antimicrobial agents (e.g. risin), substances for use in foodstuffs (e.g. in sausages, cheese) metabolic enzymes, vitamins (e.g. B12), uroporphyrinogen

(III) methyltransferase (UP III MT), *cob*A, antigens and therapeutic agents. As will be seen, the hosts can produce a wide variety of substances, not just polypeptides, which may be either the desired product or may be used to produce the desired product.

The heterologous gene may have a therapeutic effect on a human or animal. Such a gene may comprise an antigen, for example from a pathogenic organism. The host, such as Propionibacterium, comprising a polynucleotide with such a heterologous gene may be used act as or in a vaccine, and may provide protection against the pathogens.

The heterologous antigen may be a complete protein or a part of a protein containing an epitope. The antigen may be from a bacterium, a virus, a yeast or a fungus.

The host cell forms the third aspect of the invention and comprises a polynucleotide or vector of the first or second aspect. The host cell may be a bacterium e.g. of Actinomycetes. The bacterium may be a Propionibacterium or *E. coli*. The Propionibacterium may be P. freudenreichii, P. jensenii or P. acidipropionici.

In the fourth aspect the invention provides a process for producing a host cell of the third aspect, the process comprising transforming or transfecting a host cell with a polynucleotide or vector of the first or second aspect. Suitable transformation techniques can be found in Sambrook *et al*, 1989.

In a fifth aspect the invention provides a process for the preparation of a polypeptide encoded by the polynucleotide or vector of invention present in host cell of the invention comprising placing the host cell in conditions where expression of the polypeptide occurs.

This aspect of the invention thus provides a process for the preparation of a polypeptide encoded by a given gene, which process comprises cultivating a host cell transformed or transfected with an expression vector

comprising the gene, under conditions to provide for an expression of the said polypeptide, and optionally recovering the expressed polypeptide. The host cell may be of the class Actinomycetes, or a gram positive bacteria such as Propionibacterium or *E. coli*.

Promoters, elongation factor genes, ribosomal RNA, antibiotic resistance genes or synthetic promoters (e.g. designed on consensus sequences) and other expression regulation signals present in the polynucleotide or vector can be those which are compatible with expression in the host cell. Such promoters include the promoters of the endogenous genes of the host cell.

Culturing conditions may be aerobic or anaerobic conditions, depending on the host. For a fermentation process the host cell would be placed in anaerobic, and then possibly aerobic, conditions. The compound produced, such as an expressed polypeptide, may then be recovered, e.g. from the host cell or fermentation medium. The expressed polypeptide may be secreted from the host cell. Alternatively the polypeptide may not be secreted from the host cell. In such a case the polypeptide may be expressed on the surface of the host cell. This may be desirable, for example, if the polypeptide comprises an antigen to which an immune response is desired in human or animal.

A homologous gene that may be present in the vector of the invention may be *cob*A. A host cell comprising this vector may therefore be capable of producing a compound such as vitamin B12 from a substrate or the compound may be the product of an enzyme. The invention specifically provides a process for the preparation of vitamin B12 comprising cultivating or fermenting such a host cell under conditions in which the UP(III) MT gene is expressed. The expressed enzyme can be contacted with a suitable substrate under conditions in which the substrate is converted to vitamin B12.

As described above the polynucleotide of the

invention may comprise a heterologous gene which is a therapeutic gene. Thus the invention includes a host cell comprising a vector of the invention which comprises a therapeutic gene for use in a method of treatment of the human or animal body by therapy. Such a host cell may be Propionibacterium. The host cell may be alive or dead.

The host cell can be formulated for clinical administration by mixing them with a pharmaceutically acceptable carrier or diluent. For example they can be formulated for topical, parenteral, intravenous, intramuscular, subcutaneous, oral or transdermal administration. The host cell may be mixed with any vehicle which is pharmaceutically acceptable and appropriate for the desired route of administration. The pharmaceutically acceptable carrier or diluent for injection may be, for example, a sterile or isotonic solution such as Water for Injection or physiological saline.

The dose of the host cells may be adjusted according to various parameters, especially according to the type of the host cells used, the age, weight and condition of the patient to be treated; the mode of administration used; the condition to be treated; and the required clinical regimen. As a guide, the number of host cells administered, for example by oral administration, is from 10^7 to 10^{11} host cells per dose for a 70 kg adult human.

The routes of administration and dosages described are intended only as a guide since a skilled practitioner will be able to determine radically the optimum route of administration and dosage of any particular patient and condition.

A sixth aspect of the invention provides a polypeptide of the invention comprising one of the amino acid sequences set out in SEQ ID NO: 2 or 3 or a substantially homologous sequence, or of a fragment of either of these sequences. The polypeptide may be one

encoded by a polynucleotide of the first aspect. In general, the naturally occurring amino acid sequences shown in SEQ ID NO: 2 or 3 are preferred. However, the polypeptides of the invention include homologues of the natural sequences, and fragments of the natural sequences and their homologues, which have the activity of the naturally occurring polypeptides. One such activity may be to effect the replication of the polynucleotide of the invention. In particular, a polypeptide of the invention may comprise:

(a) the protein of SEQ ID No: 2 or 3; or
(b) a homologue thereof from Actinomycetes, such as Propionibacterium freudenreichii or other Propionibacterium strains; or
(c) a protein at least 70% homologous to (a) or (b).

A homologue may occur naturally in a Propionibacterium and may function in a substantially similar manner to a polypeptide of SEQ ID NO: 2 or 3. Such a homologue may occur in Actinomycetes or gram positive bacteria.

A protein at least 70% homologous to the proteins of SEQ ID NO: 2 or 3 or a homologue thereof will be preferably at least 80 or 90% and more preferably at least 95%, 97% or 99% homologous thereto over a region of at least 20, preferably at least 30, for instance at least 40, 60 or 100 or more contiguous amino acids. Methods of measuring protein homology are well known in the art and it will be understood by those of skill in the art that in the present context, homology is calculated on the basis of amino acid identity (sometimes referred to as "hard homology").

The sequences of the proteins of SEQ ID NO: 2 and 3 and of homologues can thus be modified to provide other polypeptides within the invention.

Amino acid substitutions may be made, for example from 1, 2 or 3 to 10, 20 or 30 substitutions. The modified polypeptide generally retains activity its natural activity. Conservative substitutions may be made, for example according to the following Table. Amino acids in the same block in the second column and preferably in the same line in the third column may be substituted for each other:

ALIPHATIC	Non-polar	G A P
		I L V
	Polar - uncharged	C S T M
		N Q
	Polar - charged	D E
		K R
AROMATIC		H F W Y

Polypeptides of the invention also include fragments of the above-mentioned full length polypeptides and variants thereof, including fragments of the sequences set out in SEQ ID NO: 2 or 3. Such fragments can retain the natural activity of the full-length polypeptide.

Suitable fragments will be at least about 5, e.g. 10, 12, 15 or 20 amino acids in size. Polypeptide fragments of SEQ ID No: X and homologues thereof may contain one or more (e.g. 2, 3, 5, or 10) substitutions, deletions or insertions, including conserved substitutions.

Polypeptides of the invention may be in a substantially isolated form. A polypeptide of the invention may also be in a substantially purified form, in which case it will generally comprise the polypeptide in a preparation in which more than 90%, e.g. 95%, 98% or 99% of the polypeptide in the preparation is a polypeptide of the invention.

A polypeptide of the invention may be labelled with

a revealing label. The revealing label may be any suitable label which allows the polypeptide to be detected. Suitable labels include radioisotopes, e.g. ^{125}I, ^{35}S, enzymes, antibodies, polynucleotides and linkers such as biotin.

As will apparent from the discussion, the host cells of the third aspect can be used to produce not only the recombinant proteins, but also other compounds of interest, including non-proteins such as inorganic chemicals, in particular vitamins. A seventh aspect of the present invention therefore relates to a process for the production of a compound, the process comprising culturing or fermenting host cells of the third aspect under conditions whereby the desired compound is produced. Although this compound may be a polypeptide, for example a polypeptide of the second aspect, it may also be one of the compounds mentioned in the previous discussion concerning genes to be expressed. Clearly inorganic compounds will not be expressed by a gene, but they may be produced by an enzyme, or the polypeptide or enzyme may assist the host cell in the production of the desired compound. These compounds may be produced inside the cell, and later isolated, for example following lysis of the host cell, or they may pass through the wall of the host cell into a surrounding medium, which may be a fermentation medium, for example an aqueous solution. In this way, the host cells can be cultured in an aqueous medium that comprises cells and nutrients for the cells, for example a assimilable sources of carbon and/or nitrogen.

The invention additionally encompasses the compound produced by this process, whether or not it is recombinant polypeptide. Compounds specifically contemplated are vitamins, such as vitamin B12 (cobalamin).

In some cases the compound need not be isolated either from the fermentation medium or from the host

cells. The host cells may themselves be used in particular applications, for example in, or in the manufacturing, of foodstuffs such as sausages, or in cheese making, or the host cells may for example be included in an animal feed, such as when the host cells contain a compound to be ingested by the animal in question. The invention therefore extends to the use of these compounds or the host cells, in the production of foodstuffs such as cheeses and sausages. The invention also in contemplates foodstuffs or animal feed comprising host cells or a compound in the invention.

The fermentation may have one or two phases or stages. These may be for example a growth and/or production phase, or anaerobic and/or aerobic phase. Preferably, there will be a growth and/or anaerobic phase, and suitably also (e.g. afterwards) a production and/or aerobic phase.

Both the carbon and/or nitrogen sources may be complex sources or individual compounds. For carbon, it is preferred that this is glucose. For nitrogen, appropriate sources include yeast extract or ammonia or ammonium ions.

Preferred features and characteristics of one aspect of the invention are suitable for another aspect mutatis mutandis.

The invention is illustrated by the accompanying drawings in which:

Figure 1 is a restriction map of a vector within the invention, p545 obtained from *P. freudenreichii* LMG 16545 (CBS 101022);

Figures 2a and 2b each contain two vector maps of two vectors, all four vectors being within the invention;and

Figures 3 and 4 show two open reading frames of p545 of P. freudenreichii CBS 101022, respectively. The numbering of the nucleotides in these figures is arbitrary and does not relate to the numbering in SEQ ID

NO:1.

The invention will now be described, by way of example, by reference to the following Examples, which are not to be construed as being limiting.

Example 1

Screening of *Propionibacterium* strains

A collection of 75 nonpathogenic strains of Propionibacterium was screened for the presence of indigenous plasmids. The majority of strains were obtained from the BCCM/LMG culture collection (Ghent, Belgium), although some strains were obtained from ATCC (Rockville, Md., USA) or from DSM (Braunschweig, Germany). Screening was performed using a small scale plasmid isolation procedure. First bacteria were cultivated anaerobically in MRS medium (DeMan *et al.*, 1960) for 48 hrs at 30 °C. Plasmids were then purified from the bacteria using a plasmid DNA isolation procedure originally developed for *E. coli* (Birnboim and Doly, 1979) with some modifications: cells from a 5 ml culture were washed in a 25% sucrose, 50 mM Tris-HCl pH8 solution, resuspended in 250µl TENS (25% sucrose + 50mM NaCl + 50 mM Tris-HCl + 5mM EDTA pH8), containing 10mg/ml lysozyme (Boehringer Mannheim), and incubated at 37°C for 20-30 minutes. The bacterial cells were then lysed in 500µl of 0.2 N NaOH/1% SDS (2-5 minute incubation on ice). After addition of 400µl 3M NaAc pH4.8 (5 minutes on ice) and subsequent extraction with phenol/chloroform, the DNA was precipitated by addition of isopropanol.

The DNA was analysed by electrophoresis on 1% agarose gels, and visualised by ethidium bromide. Whereas most strains were negative, i.e. did not reveal the presence of indigenous plasmids in this analysis, the majority of strains that proved positive contained large (≥20 kb) plasmids. Smaller plasmids were observed in 6 strains. Of these, *P. jensenii* LMG16453, *P. acidipropionici* ATCC4875, *P. acidipropionici* LMG16447 and

a nonspecified *Propionibacterium* strain (LMG16550) contained a plasmid in the size range of 6-10 kb. Two strains (*P. freudenreichii* LMG16545 and *P. freudenreichii* LMG16546) showed an identical plasmid profile of 3, possibly 4 plasmids. Of the two smallest plasmids, the one most abundantly present had an estimated size of 3.6 kb. The 3.6 kb plasmids from LMG16545 and LMG16546 were chosen for further analysis.

Example 2

Analysis of an indigenous plasmid from strains LMG16545 and LMG16546

The 3.6 kb plasmids were isolated from both strains and further purified by CsCl-ethidium bromide density gradient ultracentrifugation (Sambrook *et al.*, 1989). Limited restriction maps were made of both preparations and these turned out to be identical (Sambrook *et al.*, 1989). The restriction map of the 3.6 kb plasmid is shown in Fig.1. Restriction enzymes and T4 ligose were obtained from New England Biolabs or GIBCO BRL.

The 3.6 kb plasmid from strain LMG16545 (from here on referred to as p545) was radioactively labeled and used in Southern blot hybridization experiments. Hybridisation conditions were 0.2 x SSC, 65°C.It reacted equally well with both LMG16545 and LMG16546 plasmid DNA extracts, supporting the close relationship of these strains, whereas a plasmid DNA extract from *P.acidipropionici* ATCC4875, that harbors a 4.9 kb plasmid called pTY1 or pRG01 (Rehberger and Glatz, 1990), failed to react.

The DNA sequence of plasmid p545 was determined with fluorescent dye labeled dideoxyribonucleotides in an Applied Biosystems 373A automatic sequencer, and is included as SEQ ID No: 1 in the sequence listing. Sequence analysis was performed on plasmid DNA that had been linearized with *Eco*RI and inserted into *Eco*RI digested pBluescript SKII+ DNA (Stratagene, La Jolla,

Ca., USA). Computer assisted analysis of the sequence thus obtained using BLAST search (Altschul *et al*, 1990) revealed homologies to proteins involved in replication of plasmids from several GC-rich organisms (e.g., pAL5000 encoded *rep*A and *rep*B from *Mycobacterium fortuitum* [see for instance Labidi *et al*, 1992; Stolt and Stoker, 1996] show 28-30% identity and 34-38% similarity with the respective putative replication proteins from plasmid p545; pXZ10142 from *Corynebacterium glutamicum* [PIR Accession Number S32701] is another example of plasmids encoding replication proteins homologous to the p545 putative replication proteins). Database comprisons with homologous sequences are shown in Examples 7 and 8.

Example 3

Construction of *E. coli*/*Propionibacterium* shuttle vectors

E. coli plasmid pBR322 was digested with *Eco*RI and *Ava*I and the smaller fragment thus generated -measuring 1.4 kb and encompassing the tetracyclin resistance conferring gene- was replaced by a synthetic duplex DNA. The synthetic duplex DNA was designed so as to link *Eco*RI and *Ava*I ends and to supply a number of unique restriction enzyme recognition sites (SEQ ID NO:4):

5' -
AATTCAAGCTTGTCGACGTTAACCTGCAGGCATGCGGATCCGGTACCGATATCAGAT
CT - 3'
3' -
GTTCGAACAGCTGCAATTGGACGTCCGTACGCCTAGGCCATGGCTATAGTCTAGAAG
CC - 5'

The following restriction enzyme recognition sites are supplied in this way:
*Eco*RI (restored), *Hin*dIII, *Sal*I, *Hpa*I, *Pst*I, *Sph*I, *Bam*HI, *Acc*65I, *Eco*RV, *Bgl*II (*Ava*I is not restored).

This synthetic DNA was ligated to the large fragment and the ligation mixture transferred back to *E. coli* (T4

ligase was used). A plasmid of the expected composition was obtained. It was named pBR322ΔI. The multiple cloning site can be used to introduce a selection marker as well as plasmid p545 DNA.

As an example the construction of an *E. coli/ Propionibacterium* shuttle plasmid conferring resistance to erythromycin will be described.

A 1.7 kb *Acc*65I fragment from the *Saccharopolyspora erythraea* NRRL2338 erythromycin biosynthesis cluster and containing the erythromycin resistance conferring gene (Thompson *et al.*, 1982; Uchijama and Weisblum, 1985; Bibb *et al.*, 1985) was inserted into *Acc*65I linearized pBR322ΔI. Then the newly derived construct, named pBRES, was linearized with *Eco*RVand ligated to p545 DNA that had been digested with *Bsa*BI. *E.coli* transformants were found to harbor a vector with the correct insert, in both orientations.The resulting plasmid vectors were named pBRESP36B1 and pBRESP36B2 (Fig. 2a and 2b).

Plasmid vector constructs were also obtained with p545 DNA linearized in an other restriction site situated outside the putative replication region, namely *Alw*NI. For this construction the pBRES vector had to be provided with a suitable cloning site. An adaptor was designed consisting of two complementary oligonucleotides of the following composition (SEQ ID NO's 6 and 7):

5' GTACCGGCCGCTGCGGCCAAGCTT 3'
5' GATCAAGCTTGGCCGCAGCGGCCG 3'

Annealing of these oligo's creates a double stranded DNA fragment with *Acc*65I and *Bgl*II cohesive ends respectively, which moreover contains an internal *Sfi*I restriction site, that provides ends compatible to the *Alw*NI digested p545 plasmid. This adaptor was cloned in pBRES between the *Bgl*II and the proximal *Acc*65I site. The pBRES-Sfi vector thus obtained was subsequently digested by *Sfi*I and ligated to *Alw*NI digested p545.

Transformation of E.coli yielded transformants with the correct vector as confirmed by restriction enzyme analysis. The vector obtained was named pBRESP36A (Fig.2).

Example 4

Transformation of *Propionibacterium* with *E. coli/ Propionibacterium* shuttle vectors

Transformation of *Propionibacterium freudenreichii* strain ATCC6207 with pBRESP36B1 will be described.

The bacterial cells are cultivated in SLB (sodium lactate broth; de Vries *et al.*, 1972) at 30°C to a stationary growth phase, and subsequently diluted 1:50 in fresh SLB. After incubation at 30°C for around 20 hours, cells (now in the exponential growth phase) are harvested and washed extensively in cold 0.5M sucrose. Subsequently cells are washed once in the electroporation buffer, consisting of 0.5M sucrose, buffered by 1mM potassium-acetate, pH5.5, and finally resuspended in this electroporation buffer in about 1/100 of the original culture volume. Cells are kept on ice during the whole procedure.

For the electroporation (apparatus from BIORAD), 80 - 100 µl of cell suspension is mixed with ±1 µg of DNA (or smaller amounts), in a cooled 1 or 2 mm electroporation cuvette, and an electric pulse is delivered. Optimal pulse conditions were found to be 25kV/cm at 200 Ω resistance and 25µF capacitance. However, lower and higher voltages (also at 100Ω) also yield transformants.

Immediately after the pulse, 900 µl cold SLB is added to the pulsed cell suspension and these are subsequently incubated for 2.5 to 3 hours at 30°C before plating appropriate dilutions on SLB/agar plates containing 10µg/ml erythromycin. After a 5 to 7 day incubation period at 30°C under anaerobic conditions, transformants were detected.

DNA isolated from *E. coli* DH5α (PROMEGA) yields a transformation efficiency of 20 - 30 transformants per μg DNA. A 10-100 fold higher efficiency is achieved when DNA is isolated from *E. coli* JM110 (dam⁻, dcm⁻ strain). *E. coli* transformation was done according to BIORAD instructions.

Transformants contained the authentic vectors, indistinguishable from the original plasmid DNA used for transformation of ATCC6207. This was shown by restriction enzyme analysis of plasmid DNA isolated from the transformants by the small scale plasmid DNA isolation procedure refered to before.

Vectors were exclusively present as autonomously replicating plasmids. Southern blot hybridization (Southern, 1975) with total DNA isolates showed that chromosomal DNA did not hybridise to the vector DNA used as a probe, indicating that no chromosomal integration of plasmid DNA occurs.

Transformation was also successful with vectors pBRESP36B2 and pBRESP36A, indicating that functionality of the vector was independent of the orientation of p545 or the cloning site used. Also in this case the authenticity of the vectors was confirmed.

Moreover, transformation of *P. freudenreichii* strain ATCC6207 with DNA isolated from a *Propionibacterium* transformant resulted in a 10^5-10^6 fold increased transformation efficiency as compared to that obtained with DNA isolated from *E. coli* DH5α.

Transformation of another *P.freudenreichii* strain, LMG16545 (the same strain from which the p545 plasmid was obtained), resulted in a transformation efficiency comparable to that of the afore mentioned ATCC strain.

Example 5

Construction of plasmid vect r containing the *cob*A gene

The construction and application of a plasmid vector aimed to increase the level of vitamin B12 (cobalamin)

synthesis in *P. freudenreichii* strain ATCC6207 will be described.

The promoter region of the gene conferring erythromycin resistance in *Saccharopolyspora erythraea* (Bibb *et al.*, 1985; Bibb *et al.*, 1994), was generated by PCR using the following primers (SEQ ID NO's 8 and 9):

forward primer: (5' - 3')
AAACTGCAGCTGCTGGCTTGCGCCCGATGCTAGTC

reverse primer: (5' - 3')
AAACTGCAGCAGCTGGGCAGGCCGCTGGACGGCCTGCCCTCGAGCTCGTCTAGAATG
TGCTGCCGATCCTGGTTGC

The PCR fragment thus generated contains an *Alw*NI site at the 5' end followed by the authentic promoter region and the first 19 amino acids of the coding region of the erythromycin gene, to ensure proper transcription and translation initiation. At the 3' end *Xba*I and *Xho*I sites are provided (for insertion of the *cob*A gene in a later stage), a terminator sequence as present downstream from the erythromycin gene, and an *Alw*NI site.

The PCR product was digested by *Alw*NI and ligated to pBRESP36B2, partially digested with *Alw*NI. Of the two *Alw*NI sites present in pBRESP36B2, only the one present in the p545 specific part of the vector will accommodate the fragment. *E. coli* transformants were obtained harboring the expected construct, named pBRES36pEt. This vector was used for further constructions as described below.

The coding sequence of *cob*A, the gene encoding uroporphyrinogen III methyltransferase, was generated by PCR from *Propionibacterium freudenreichii* strain ATCC6207, using the following primers (SEQ ID NO's 10 and 11):
forward: (5'- 3')
CTAGTCTAGACACCGATGAGGAAACCCGATGA

reverse: (5'- 3')
CCCAAGCTTCTCGAGTCAGTGGTCGCTGGGCGCGCG

The *cob*A gene thus amplified carries an *Xba*I site at the N terminal coding region, and *Hin*dIII and *Xho*I sites at the C terminal coding region.

The functionality of this *cob*A gene was confirmed by cloning the PCR product as an *Xba*I -*Hin*dIII fragment in pUC18, and subsequent transformation of *E.coli* strain JM109. Transformants with a functional *cob*A gene show a bright red fluorescence when illuminated with UV light. Plasmid DNA isolated from such a transformant was digested with *Xba*I and *Xho*I, ligated to likewise digested pBRES36pEt DNA and used for transformation of *E. coli*. DNA from several transformants was analysed by restriction enzyme digestion and gel electrophoresis. Transformants were found to bear the correct insert in the expression vector. This new vector was named pBRES36COB. This vector was subsequently transferred to *P. freudenreichii* ATCC6207 following the protocol described before. Ten of the transformants obtained were analysed and were found to harbor the pBRES36COB vector, which was again indistinguishable from the original vector used for transformation, as shown by analysis of the restriction enzyme profile. In these ten transformants the level of vitamin B12 synthesis was determined as follows:

Frozen cultures of *Propionibacterium* transformants 1 through 10, as well as a control strain containing only the vector plasmid pBRES36pEt, were inoculated in 100 ml flasks containing 50 ml of BHI (Brain Heart Infusion) / medium (Difco) and incubated for 72 hrs at 28°C without shaking. From this preculture 4 ml were transferred to 200 ml of production medium consisting of Difco yeast extract 15 g/l, Na-lactate 30 g/l, KH_2PO_4 0.5g/l, $MnSO_4$ 0.01 g/l, and $CoCl_2$ 0.005 g/l in a 500 ml shake flask and incubated at 28°C for 56 hrs without shaking, followed by

48 hrs in a New Brunswick rotary shaker at 200 rpm.

Vitamin B12 titres were measured using the HPLC method as published by Blanche (Analytical Biochemistry, 1990). Nine out of 10 transformants showed an approx. 25% higher vitamin B12 production than the control strain.

Example 6

Stability of the plasmids

All three shuttle vectors pBRESP36A, pBRESP36B1, and pBRESP36B2 are stably maintained over 30 generations of culturing of the respective transformants: no loss of Erythromycin resistence was observed as determined by viability counts on selective (erythromycin containing) and no-selective agar plates. The structural stability of the plasmid in the transformant population after 30 generations was established by plasmid DNA isolation and characterisation by restriction enzyme mapping as described above: only restriction fragments similar to those of the authentic plasmid were observed.

Example 7

Database sequence homology analysis for predicted polypeptide encoded by the first open reading frame

```
FILE NAME   : pORF1.AMI
SEQUENCE    : 227 AA
RANGE       : 1 - 227
CUTOFF      : 45                KTUP    : 2
Target      : NBRF-PIR, Release : PIR R52.0 March, 1997
Group Name : All Entry
 No. Target file  Definition                                   Match% Over. INIT
OPT
   1 JS0052                                                     37.1  194   167
292
                      10         20         30         40         50
 PORF1.AMI      MDSFETLFPESWLPRKPLASAE-KSGAYRHVTRQRALELPYIEANPLVMQSLVITDRDAS
                       : . :::  :::::..  .: ::. .:..::. .:::::: ....:...: : .
 JS0052         VSHVADEFEQLWLPYWPLASDDLLEGIYRQ-SRASALGRRYIEANPTALANLLVVDVDHP
                30         40         50         60         70         80
                60         70         80         90        100        110
 PORF1.AMI      DAD-WA-ADLAGLPSPSYVSMNRVTTTGHIVYALKNPVCLTDAARRRPINLLARVEQGLC
                ::. .: .. .. : :. .  X:... .: :.:::..::  :. :::.:.. .:  ..::
 JS0052         DAALRALSARGSHPLPNAIVGNRANGHAHAVWALNAPVPRTEYARRKPLAYMAACAEGLR
                 90         100        110        120        130        140
                120        130        140        150        160        170
 PORF1.AMI      DVLGGDASYGHRITKNPLSTAHATLWGPADALYELRALAHTLDEIHALPE-AGNPRRNVT
                 ...:: ::.  .:::::  .: .: : ..: ::.:. .. .:..  . :    ..  ...
 JS0052         RAVDGDRSYSGLMTKNPGHIAWETEWLHSD-LYTLSHIEAELGANMPPPRWRQQTTYKAA
                150        160        170        180        190        200
                 180        190        200        210        220
```

```
PORF1.AMI      RSTVGRNVTLFDTTRMWAYRAVRHSWGGPVAEWEHTVFEHIHLLNETIIAD
               ...::: .:::..:.:::X.
JS0052         PTPLGRNCALFDSVRLWAYRPALMRIYLPTRNVDGLGRAIYAECHARNAEF
                210       220       230       240       250
```

JS0052 is Mycobacterium fortuitum plasmid pAL5000

```
 No. Target file  Definition                                  Match% Over. INIT
OPT
   2 S32701                                                    32.0  125   116
186
                    50        60        70        80        90        100
 PORF1.AMI      PLVMQSLVITDRDASDADWAADLAGLPSPSYVSMNRVTTTGHIVYALKNPVCLTDAARRR

 S32701         MPSRISWSSTSTSRTHSCVRCGTETAGGLTPWLKTPFKRARTRRVGARGAIYPHRVRQAQ
                         10        20        30        40        50        60
                   110       120       130       140       150       160
 PORF1.AMI      PINLLARVEQGLCDVLGGDASYGHRITKNPLSTAHATLWGPADALYELRALAHTLDEIHA
                ...  : :..X:   ..:: .:.  ::::: :: .. :  .: ::.: .:   :.:.
 S32701         ALAYAAAVTEGLRRSVDGDKGYSGLITKNPEHTAWDSHW-VTDKLYTLDELRFWLEETGF
                       70        80        90        100       110
                  170       180       190       200       210       220
 PORF1.AMI      LPEAG-NPRRNVTRSTVGRNVTLFDTTRMWAYRAVRHSWGGPVAEWEHTVFEHIHLLNET
                .: .. .  :. .. ..:::: .:::..: ::::::X .:.: .. .... .  . ::..
 S32701         MPPESWKKTRRKSPIGLGRNCALFESARSWAYREIRHHFGDP-DGLGRSIQATAQALNQE
               120       130       140       150       160       170

 PORF1.AMI      IIAD
                ....
 S32701         LFSE
               180
```

S32701 is from Corynebacterium glutamicum

```
 No. Target file  Definition                                  Match% Over. INIT
OPT
   3 S04455                                                    29.9  221    86
259
                         10        20        30        40        50
 PORF1.AMI      MDSFETLFPESWLPRKPLASAEKSGAYRHVTRQRALELPYIEAN-PLVMQSLVI-TDRDA
                      :.:. ::.::  ... . . : .... :.  ..:..: : .. . : . .::..
 S04455         MSAVLQRFREK-LPHKPYCTNDFAYGVRILPKNIAILARFIQQNQPHALYWLPFDVDRTG
                         10        20        30        40        50
                  60        70        80        90        100       110
 PORF1.AMI      SDADWAADLAGLPSPSYVSMNRVTTTGHIVYALKNPVCLTDAARRRPINLLARVEQGLCD
                .. :: .: .  :.:. . .:. .. .:..X::  ::   . .:. .... : .:.::.
 S04455         ASIDW-SDRNC-PAPNITVKNPRNGHAHLLYALALPVRTAPDASASALRYAAAIERALCE
                60        70        80        90        100       110
               120       130       140       150       160       170
 PORF1.AMI      VLGGDASYGHRITKNPLSTAHATLWGPADALYELRALAHTLDEIHA-LPEAGNPRRNVTR
                 ::.:..:.  : ::X    :.. :  ..     :. ..:::::. . :  ....::.:..
 S04455         KLGADVNYSGLICKNP---CHPE-WQEVE----WREEPYTLDELADYLDLSASARRSVDK
                120       130       140       150       160
                 180       190       200       210       220
 PORF1.AMI      S-TVGRNVTLFDTTRMWAYRAVRHSWGGPVAEWEHTVFEHIHLLNETIIAD
                . ..::::  ::...:.:::::.:..:    ..: ..:..... :...
 S04455         NYGLGRNYHLFEKVRKWAYRAIRQGW-PVFSQWLDAVIQRVEMYNASLPVP
               170       180       190       200       210
```

S04455 is from E. coli ColE2

```
 No. Target file  Definition                                  Match% Over. INIT
OPT
   4 S04456                                                    29.9  221    86
259
                         10        20        30        40        50
 PORF1.AMI      MDSFETLFPESWLPRKPLASAEKSGAYRHVTRQRALELPYIEAN-PLVMQSLVI-TDRDA
                      :.:. ::.::  ... . . : .... :.  ..:..: : ..  : . .::..
 S04456         MSAVLQRFREK-LPHKPYCTNDFAYGVRILPKNIAILARFIQQNQPHALYWLPFDVDRTG
```

```
                    10         20         30         40         50
                   60         70         80         90        100        110
PORF1.AMI          SDADWAADLAGLPSPSYVSMNRVTTTGHIVYALKNPVCLTDAARRRPINLLARVEQGLCD
                   .. :: .: .  :.:. . .:. .. .:..X::  ::  . .:. ....  : .:..::.
S04456             ASIDW-SDRNC-PAPNITVKNPRNGHAHLLYALALPVRTAPDASASALRYAAAIERALCE
                  60         70         80         90        100        110
                  120       130        140        150        160        170
PORF1.AMI          VLGGDASYGHRITKNPLSTAHATLWGPADALYELRALAHTLDEIHA-LPEAGNPRRNVTR
                    ::.:..:.  : ::X    :.. :  ..     :. ..:::::. . :  ....:.:..
S04456             KLGADVNYSGLICKNP---CHPE-WQEVE----WREEPYTLDELADYLDLSASARRSVDK
                   120        130           140           150        160
                    180        190        200        210        220
PORF1.AMI          S-TVGRNVTLFDTTRMWAYRAVRHSWGGPVAEWEHTVFEHIHLLNETIIAD
                   . ..:::  ::...:.:::::.:..:  . ..: ..:..... :...
S04456             NYGLGRNCYLFEKGRKWAYRAIRQGW-PAFSQWLDAVIQRVEMYNASLPVP
                  170        180        190        200        210
```

SO4456 is from E. coli Col E3

Example 8

Database sequence homology analysis for predicted polypeptide encoded by the second reading frame

```
FILE NAME   : pORF2.ami
SEQUENCE    : 85 AA
RANGE       : 1 - 85
CUTOFF      : 45                KTUP    : 2
Target      : NBRF-PIR, Release : PIR R52.0 March, 1997
Group Name : All Entry
 No. Target file  Definition                                    Match% Over. INIT
OPT
   1 S32702                                                      53.3   75   207
207
                    10         20         30         40         50         60
 pORF2.ami         MTTRERLPRNGYSIAAAAKKLGVSESTVKRWTSEPREEFVARVAARHARIRELRSEGQSM
                   X:.:.:.:::: .: ..:.  :.:..:..:::::.:::::.:.....:.:::::..: ::
 S32702            MTKRTRIPRNGKTIREVAEGTGLSTATIERWTSAPREDYLAQANEKRVRVQELRAKGLSM
                   20         30         40         50         60         70
                          70         80
 pORF2.ami         RAIAAEVGVSVGTVHYALNKNRTDA
                   :::::::.: ::: :X
 S32702            RAIAAEIGCSVGLVHRYVKEVEEKK
                   80         90        100
```

S32702 is from Corynebacterium glutamicum

Predicted amino acid sequence encoded by the first open reading frame

MDSFETLFPESWLPRKPLASAEKSGAYRHVTRQRALELPYIEANPLVMQSLVITDRDA
SDADWAADLAGLPSPSYVSMNRVTTTGHIVYALKNPVCLTDAARRRPINLLARVEQGL
CDVLGGDASYGHRITKNPLSTAHATLWGPADALYELRALAHTLDEIHALPEAGNPRRN
VTRSTVGRNVTLFDTTRMWAYRAVRHSWGGPVAEWEHTVFEHIHLLNETIIAD

Predicted amino acid sequence encoded by the second open reading frame

MTTRERLPRN GYSIAAAAKK LGVSESTVKR WTSEPREEFV ARVAARHARI
RELRSEGQSM RAIAAEVGVS VGTVHYALNK NRTDA

References

Altschul *et al* (1990) J. Mol. Biol. 215, 403

Bibb *et al.*, (1985) Gene 38, 215

Bibb *et al.*, (1994) Mol. Microbiol. 14(3), 533

Birnboim and Doly (1979) Nucleic Acids Res. 7, 1513

Blanche (1990) Anal. Biochem. 189, 24

DeMan *et al* (1960) J. Appl. Bacteriol. 36, 130

Labidi *et al* (1992) Plasmid 27, 130

Rehberger and Glatz (1990) Appl.Environ. Microbiol. 59, 83

Rossi *et al* (1996) Res. Microbiol. 147, 133

Sambrook *et al* (1989) Molecular cloning, Cold Spring Harbor Laboratory Press.

Sattler *et al* (1995) J. Bact. 177, 1564

Southern (1975) J. Mol. Biol. 98, 503

Stolt and Stoker (1996) Microbiol. 142, 2795

Thompson *et al* (1982) Gene 20, 51

Uchijama and Weisblum (1985) Gene 38, 103

de Vries *et al* (1972) J. Gen. Microbiol. 71, 515

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:
(A) NAME: Gist-brocades B.V.
(B) STREET: Wateringseweg 1
(C) CITY: Delft
(E) COUNTRY: The Netherlands
(F) POSTAL CODE (ZIP): 2600 MA

(ii) TITLE OF INVENTION: Propionibacterium Vector

(iii) NUMBER OF SEQUENCES: 11

(iv) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3555 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: circular

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Propionibacterium freudenreichii
(C) INDIVIDUAL ISOLATE: CBS101022 LMG16545

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 273..1184
(D) OTHER INFORMATION: /gene= "ORF1"

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1181..1438
(D) OTHER INFORMATION: /gene= "ORF2"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

GTCGACCCTG ACAGCCGGCG AGCAGTTCAG GCGAAGATCG CACAGCTGCG CGAGGAACTA 60

GCCGCAATGC CCGAACACGC CCCAGCCATC CCTTGGAGCA GGTGGCAGCG TCAGGGGAGT 120

```
CGGGGGATGT TTGGCAGGGG ATGTGGAAAG AGAGTTCGCT TTGCTCACAT GGCTCAACCG 180

GGTAACTAAC TGATATGGGG TCTTCGTCGC CCACTTTGAA CACGCCGAGG AATGGACCAC 240

GCTGAACGTG ACTCGCATGC TTCACTGCAT GT ATG GAT TCG TTC GAG ACG TTG   293
                                       Met Asp Ser Phe Glu Thr Leu
                                         1               5

TTC CCT GAG AGC TGG CTG CCA CGC AAG CCG CTG GCG TCA GCC GAG AAG   341
Phe Pro Glu Ser Trp Leu Pro Arg Lys Pro Leu Ala Ser Ala Glu Lys
          10                  15                  20

TCT GGG GCG TAC CGG CAC GTG ACT CGG CAG AGG GCG CTG GAG CTG CCT   389
Ser Gly Ala Tyr Arg His Val Thr Arg Gln Arg Ala Leu Glu Leu Pro
     25                  30                  35

TAC ATC GAA GCG AAC CCG TTG GTC ATG CAG TCC TTG GTC ATC ACC GAT   437
Tyr Ile Glu Ala Asn Pro Leu Val Met Gln Ser Leu Val Ile Thr Asp
 40                  45                  50                  55

CGA GAT GCT TCG GAT GCT GAC TGG GCC GCA GAC CTC GCT GGG CTG CCT   485
Arg Asp Ala Ser Asp Ala Asp Trp Ala Ala Asp Leu Ala Gly Leu Pro
                 60                  65                  70

TCA CCG TCC TAC GTG TCC ATG AAC CGT GTC ACG ACC ACC GGA CAC ATC   533
Ser Pro Ser Tyr Val Ser Met Asn Arg Val Thr Thr Thr Gly His Ile
             75                  80                  85

GTC TAT GCC TTG AAG AAC CCT GTG TGT CTG ACC GAT GCC GCG CGG CGA   581
Val Tyr Ala Leu Lys Asn Pro Val Cys Leu Thr Asp Ala Ala Arg Arg
          90                  95                  100

CGG CCT ATC AAC CTG CTC GCC CGC GTC GAG CAG GGC CTA TGC GAC GTT   629
Arg Pro Ile Asn Leu Leu Ala Arg Val Glu Gln Gly Leu Cys Asp Val
     105                 110                 115

CTC GGC GGC GAT GCA TCC TAC GGG CAC CGG ATC ACA AAG AAC CCG CTC   677
Leu Gly Gly Asp Ala Ser Tyr Gly His Arg Ile Thr Lys Asn Pro Leu
120                 125                 130                 135

AGC ACC GCC CAT GCG ACC CTC TGG GGC CCC GCA GAC GCG CTC TAC GAG   725
Ser Thr Ala His Ala Thr Leu Trp Gly Pro Ala Asp Ala Leu Tyr Glu
                140                 145                 150

CTG CGC GCC CTC GCA CAC ACC CTC GAC GAG ATC CAC GCA CTG CCG GAG   773
Leu Arg Ala Leu Ala His Thr Leu Asp Glu Ile His Ala Leu Pro Glu
            155                 160                 165

GCA GGG AAC CCG CGT CGC AAC GTC ACC CGA TCA ACG GTC GGC CGC AAC   821
Ala Gly Asn Pro Arg Arg Asn Val Thr Arg Ser Thr Val Gly Arg Asn
        170                 175                 180

GTC ACC CTG TTC GAC ACC ACC CGC ATG TGG GCA TAC CGG GCC GTC CGG   869
Val Thr Leu Phe Asp Thr Thr Arg Met Trp Ala Tyr Arg Ala Val Arg
    185                 190                 195

CAC TCC TGG GGC GGC CCG GTC GCC GAA TGG GAG CAC ACC GTA TTC GAG   917
```

```
His Ser Trp Gly Gly Pro Val Ala Glu Trp Glu His Thr Val Phe Glu
200                 205                 210                 215

CAC ATC CAC CTA CTG AAC GAG ACG ATC ATC GCC GAC GAA TTC GCC ACA   965
His Ile His Leu Leu Asn Glu Thr Ile Ile Ala Asp Glu Phe Ala Thr
                220                 225                 230

GGC CCC CTC GGC TTG AAC GAA CTT AAG CAC TTA TCT CGA TCC ATT TCC  1013
Gly Pro Leu Gly Leu Asn Glu Leu Lys His Leu Ser Arg Ser Ile Ser
            235                 240                 245

CGA TGG GTC TGG CGC AAC TTC ACC CCC GAA ACC TTC CGC GCA CGC CAG  1061
Arg Trp Val Trp Arg Asn Phe Thr Pro Glu Thr Phe Arg Ala Arg Gln
        250                 255                 260

AAA GCG ATC AGC CTC CGT GGA GCA TCC AAA GGC GGC AAA GAA GGC GGC  1109
Lys Ala Ile Ser Leu Arg Gly Ala Ser Lys Gly Gly Lys Glu Gly Gly
    265                 270                 275

CAC AAA GGC GGC ATT GCC AGT GGC GCA TCA CGG CGC GCC CAT ACC CGT  1157
His Lys Gly Gly Ile Ala Ser Gly Ala Ser Arg Arg Ala His Thr Arg
280                 285                 290                 295

CAA CAG TTC TTG GAG GGT CTC TCA TGACCACACG TGAACGTCTC CCCCGCAACG 1211
Gln Gln Phe Leu Glu Gly Leu Ser
                300

GCTACAGCAT CGCCGCTGCT GCGAAAAAGC TCGGTGTCTC CGAGTCCACC GTCAAGCGGT
1271

GGACTTCCGA GCCACGCGAG GAGTTCGTGG CCCGCGTTGC CGCACGCCAC GCGCGGATTC
1331

GTGAGCTCCG CTCGGAGGGT CAGAGCATGC GTGCGATTGC TGCCGAGGTC GGGGTTTCCG
1391

TGGGCACCGT GCACTACGCG CTGAACAAGA ATCGAACTGA CGCATGACCG TAACGCCGCA
1451

CGATGAGCAT TTTCTTGATC GTGCACCGCT TGGCACTACG TTCGCGTGCG GTTGCACAGT
1511

GCGCGCCACG TTCTTATCCT GCGGCCATTG TGGCTACAGC CAATGGGGGG CATCAGCAAC
1571

GGACGTTGAA CCCGGTGGGC AAGTGTTACT CAGGGGGACA TGCCCAGTCT GCGGCGCTCG
1631

GATTGACGGT ATGGCAGTCG TGCATGCGGC CCCACCGTCA AACTCATTCA GGTATCAGTG
1691

AGAACCCTCA TGGCACCCCC TCGTGACACG TTCTCGTTGC GATCAGCTGC TGTGCGTGCG
1751

GGCGTGAGCG TTTCTACGCT GCGGCGCAGG AAATCAGAGC TTGAGGCTGC CGGAGCGACG
1811
```

GTAGACCCGT CCGGTTGGGT GGTGCCACTG CGTGCACTCA AGGTCGTTTT TGGGGTGTCA
1871

GATGAGACCT CGAATGCGCC CGGTCATGAC GCTGAGTTAG TGGCGCAGCT GCGCTCTGAG
1931

AACGAGTTTT TACGGCGTCA GGTCGAGCAG CAGGCGCGCA CGATCGAACG GCAGGCTGAG
1991

GCACACGCGG TGGTCTCAGC GCAGCTCACA CGGGTTGGCC AGCTTGAGGC CGGCGACGCA
2051

GCAGCACCGA CACTGGCACC CGTTGAAAGG CCGGCTCCGC GACGGCGGTG GTGGCAGCGT
2111

CGGTAGCGGT CAGGATCGCT CTGGCGTGAC GAGTGTGTCT GGCAGTGCGA ACAGTTGCTC
2171

GACCAGTGGC AGCAGAAGCG AGATCGCTGC GTGGTGCTGT TCCTCGGTCA GTTCGTCGAG
2231

GACTGGCGGG TCTTGCTGCG TCCAGCCGAT CGCCTCGGCG GCCAAGGTCA GTTCCAAGCT
2291

GTGCCAACGC ACACGCCCCT CGGCTGACAG CTGAGTCTCG AACTGTGCAA CTGGACCGGC
2351

CGGAAGATGC ACGTTGCCGA GGTCGTGAGT GGCCAAGCGC ACGTCAAAGA GTGCTGCTTC
2411

GTAGCCGCGC AGAAATGGCA GTGCTCGGTC GATTCGGATC GGCCTGCCCA GGTACATTCC
2471

GGGCCGCTTG ATGAACGCCT CCGCGTAGAA GCGCACCGTT CTCGGCCCGG CCTCGTGATC
2531

TGTCACTGTG CACGCTCCTC TCGATGGTTC TCGACGCTAC CGGAGACCAC CGACGTTCAT
2591

GCCCAGCGCA GCGACCTGAA AGGACCAAGC CGAGTTAGCC GTGCTAACCG TATAGCTTGC
2651

TCCGTCGCCT CTGAGGGCAA CCACCTGCGC AGCAGGTGGG CGGCAGCCCG CGCGCAAGCG
2711

CCTACCGGGT TTGGGCACAG CCCATAAATC AACGCCTCCG GTGTTGAAGC GATCGTGTGT
2771

CACGATTGCT ATGCTTGCTA CCCCTTCAGG GTTTTCGTAT ACACAAATCA AGTTTTTTCG
2831

TATACGCTAA TGCCATGAGT GAGCATCTAC TGCACGGCAA GCCCGTCACC AACGAGCAGA
2891

TTCAGGCATG GGCAGACGAG GCCGAGGCCG GATACGACCT GCCCAAACTC CCCAAGCCAC
2951

GGCGCGGACG CCCGCCCGTA GGAGACGGTC CGGGCACCGT CGTACCCGTG CGTCTCGACG
3011

CGGCCACCGT TGCCGCTCTC ACAGAACGAG CAACAGCCGA GGGCATCACG AACCGTTCAG
3071

ACGCGATCCG AGCCGCAGTC CACGAGTGGA CACGGGTTGC CTGACCTCCA CGACTCAGCA
3131

CGCAAGCACT ACCAACGAGA CCGGCTCGAC GACACGGCCG TGCTCTACGC GGCCACCCAC
3191

GTTCTCAACT CCCGGCCACT CGACGACGAA GACGACCCGC GCCGCTGGCT CATGATCGGA
3251

ACCGACCCAG CAGGCCGCCT ACTCGAACTC GTCGCACTGA TCTACGACGA CGGCTACGAA
3311

CTGATCATCC ACGCAATGAA AGCCCGCACC CAATACCTCG ACCAGCTCTA ACCAAGAAAG
3371

GAACCTGATG AGCGACCAGC TAGACAGCGA CCGCAACTAC GACCCGATGA TCTTCGACGT
3431

GATGCGCGAG ACCGCGAACC GCGTCGTCGC CACGTACGTT GCATGGGAAG ATGAAGCCGC
3491

TGATCCCCGC GAGGCTGCGC ACTGGCAGGC CGAGCGATTC CGCACCCGGC ACGAGGTGCG
3551

CGCC
3555

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 303 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Met Asp Ser Phe Glu Thr Leu Phe Pro Glu Ser Trp Leu Pro Arg Lys
1 5 10 15

Pro Leu Ala Ser Ala Glu Lys Ser Gly Ala Tyr Arg His Val Thr Arg
20 25 30

Gln Arg Ala Leu Glu Leu Pro Tyr Ile Glu Ala Asn Pro Leu Val Met
35 40 45

Gln Ser Leu Val Ile Thr Asp Arg Asp Ala Ser Asp Ala Asp Trp Ala
50 55 60

```
Ala Asp Leu Ala Gly Leu Pro Ser Pro Ser Tyr Val Ser Met Asn Arg
 65                  70                  75                  80

Val Thr Thr Thr Gly His Ile Val Tyr Ala Leu Lys Asn Pro Val Cys
                 85                  90                  95

Leu Thr Asp Ala Ala Arg Arg Arg Pro Ile Asn Leu Leu Ala Arg Val
            100                 105                 110

Glu Gln Gly Leu Cys Asp Val Leu Gly Gly Asp Ala Ser Tyr Gly His
        115                 120                 125

Arg Ile Thr Lys Asn Pro Leu Ser Thr Ala His Ala Thr Leu Trp Gly
    130                 135                 140

Pro Ala Asp Ala Leu Tyr Glu Leu Arg Ala Leu Ala His Thr Leu Asp
145                 150                 155                 160

Glu Ile His Ala Leu Pro Glu Ala Gly Asn Pro Arg Arg Asn Val Thr
                165                 170                 175

Arg Ser Thr Val Gly Arg Asn Val Thr Leu Phe Asp Thr Thr Arg Met
            180                 185                 190

Trp Ala Tyr Arg Ala Val Arg His Ser Trp Gly Gly Pro Val Ala Glu
        195                 200                 205

Trp Glu His Thr Val Phe Glu His Ile His Leu Leu Asn Glu Thr Ile
    210                 215                 220

Ile Ala Asp Glu Phe Ala Thr Gly Pro Leu Gly Leu Asn Glu Leu Lys
225                 230                 235                 240

His Leu Ser Arg Ser Ile Ser Arg Trp Val Trp Arg Asn Phe Thr Pro
                245                 250                 255

Glu Thr Phe Arg Ala Arg Gln Lys Ala Ile Ser Leu Arg Gly Ala Ser
            260                 265                 270

Lys Gly Gly Lys Glu Gly Gly His Lys Gly Gly Ile Ala Ser Gly Ala
        275                 280                 285

Ser Arg Arg Ala His Thr Arg Gln Gln Phe Leu Glu Gly Leu Ser
    290                 295                 300
```

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 85 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

Met Thr Thr Arg Glu Arg Leu Pro Arg Asn Gly Tyr Ser Ile Ala Ala
1 5 10 15

Ala Ala Lys Lys Leu Gly Val Ser Glu Ser Thr Val Lys Arg Trp Thr
20 25 30

Ser Glu Pro Arg Glu Glu Phe Val Ala Arg Val Ala Ala Arg His Ala
35 40 45

Arg Ile Arg Glu Leu Arg Ser Glu Gly Gln Ser Met Arg Ala Ile Ala
50 55 60

Ala Glu Val Gly Val Ser Val Gly Thr Val His Tyr Ala Leu Asn Lys
65 70 75 80

Asn Arg Thr Asp Ala
85

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 59 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

AATTCAAGCT TGTCGACGTT AACCTGCAGG CATGCGGATC CGGTACCGAT ATCAGATCT
59

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 59 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

CCGAAGATCT GATATCGGTA CCGGATCCGC ATGCCTGCAG GTTAACGTCG ACAAGCTTG
59

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

GTACCGGCCG CTGCGGCCAA GCTT
24

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

GATCAAGCTT GGCCGCAGCG GCCG
24

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 35 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

AAACTGCAGC TGCTGGCTTG CGCCCGATGC TAGTC
35

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 76 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

AAACTGCAGC AGCTGGGCAG GCCGCTGGAC GGCCTGCCCT CGAGCTCGTC TAGAATGTGC
60

TGCCGATCCT GGTTGC
76

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

CTAGTCTAGA CACCGATGAG GAAACCCGAT GA
32

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 36 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

CCCAAGCTTC TCGAGTCAGT GGTCGCTGGG CGCGCG
36

CLAIMS

1. A polynucleotide comprising a sequence capable of hybridising selectively to

(i) SEQ ID NO: 1 or the complement thereof;

(ii) a sequence from the 3.6 kb plasmid of Propionibacterium freudenreichii CBS 101022; or

(iii) a sequence from the 3.6 kb plasmid of Propionibacterium freudenreichii CBS 101023.

2. A polynucleotide which a autonomously replicating plasmid that can remain extrachromosomal inside a host cell, which plasmid is derived from an endogenous Propionibacterium plasmid, and when comprising a heterologous gene is capable of expressing that gene inside the host cell.

3. A polynucleotide according to claim 1 which is autonomously replicating in a host cell.

4. A polynucleotide according to claim 3 in which the host cell is a Propionibacterium.

5. A polynucleotide according to claim 4 in which the Propionibacterium is Propionibacterium freudenreichii.

6. A polynucleotide according to any one of the preceding claims which is capable of selectively hybridising to one or more sequence(s) in SEQ ID No:1 which is (or are) necessary for autonomous replication in a Propionibacterium.

7. A polynucleotide comprising

(i) the sequence of SEQ ID No: 1 or the complement thereof;

(ii) a sequence from the 3.6 kb plasmid of Propionibacterium freudenreichii CBS 101022, or

(iii) a sequence from the 3.6 kb plasmid of Propionibacterium freudenreichii CBS 101023.

8. A vector which comprises a polynucleotide according to any one the preceding claims.

9. A vector according to claim 8 which is a

plasmid.

10. A vector according to claim 8 or 9 which additionally comprises a selectable marker.

11. A vector according to any one of claims 8 to 10 which is autonomously replicating in *E. coli*.

12. A vector according to any one of claims 8 to 11 which is an expression vector.

13. A vector according to claim 12 which comprises an endogenous gene of a Propionibacterium or a heterologous gene operatively linked to a control sequence which is capable of providing for expression of the gene.

14. A vector according to claim 13 in which the heterologous gene is the *cob*A gene.

15. A vector according to claim 13 in which the heterologous gene encodes a polypeptide which is therapeutic in a human or animal.

16. A polypeptide which comprises the sequence SEQ ID No: 2 or 3 or a sequence substantially homologous thereto, or a fragment of either said sequence, or is encoded by a polynucleotide as defined in any of claims 1 to 7.

17. A host cell comprising a polynucleotide or vector according to any one of claims 1 to 15 or which can been transformed or transfected with a vector according to any one of claims 13 to 15.

18. A host cell according to claim 17 which is a bacterium.

19. A host cell according to claim 18 which is a Propionibacterium or *E. coli*.

20. A process for producing a host cell according to any one of claims 17 to 19 comprising transforming or transfecting a host cell with a polynucleotide or vector according to any one of claims 1 to 15.

21. A process for the preparation of a polypeptide, or other compound, the process comprising cultivating or fermenting a host cell as defined in any one of claims 17

to 19 under conditions that allow expression or production of the polypeptide or compound.

22. A process according to claim 21 which is a fermentation process wherein the host cell is cultured in aerobic or anaerobic conditions.

23. A process according to claim 21 or 22 in which the expressed polypeptide or produced compound is recovered from the host cell.

24. A process according to any one of claims 22 to 23 where the polypeptide is secreted from the host cell.

25. A process according to claim 24 in which the polypeptide is expressed on the surface of the host cell.

26. A polypeptide or compound prepared by a process according to any one of claims 20 to 25.

27. A process for the production of vitamin B12 (cobalamin) comprising culturing a host cell according to any one of claims 17 to 19 under conditions in which the vitamin gene is produced and, if necessary isolating the vitamin.

28. Vitamin B12 produced in a process according to claim 27.

29. A polypeptide according to claim 26 for use in a method of treating the human or animal body by therapy.

30. A host cell according to any one of claims 17 to 19 for use in a method of treating the human or animal body by therapy or for use in an animal feed.

31. Use of a host cell according to any one of claims 17 to 19 or a polypeptide or compound according to claim 26 to make cheese or in cheesemaking, in the manufacture of a foodstuff or in an animal feed.

Figure 1

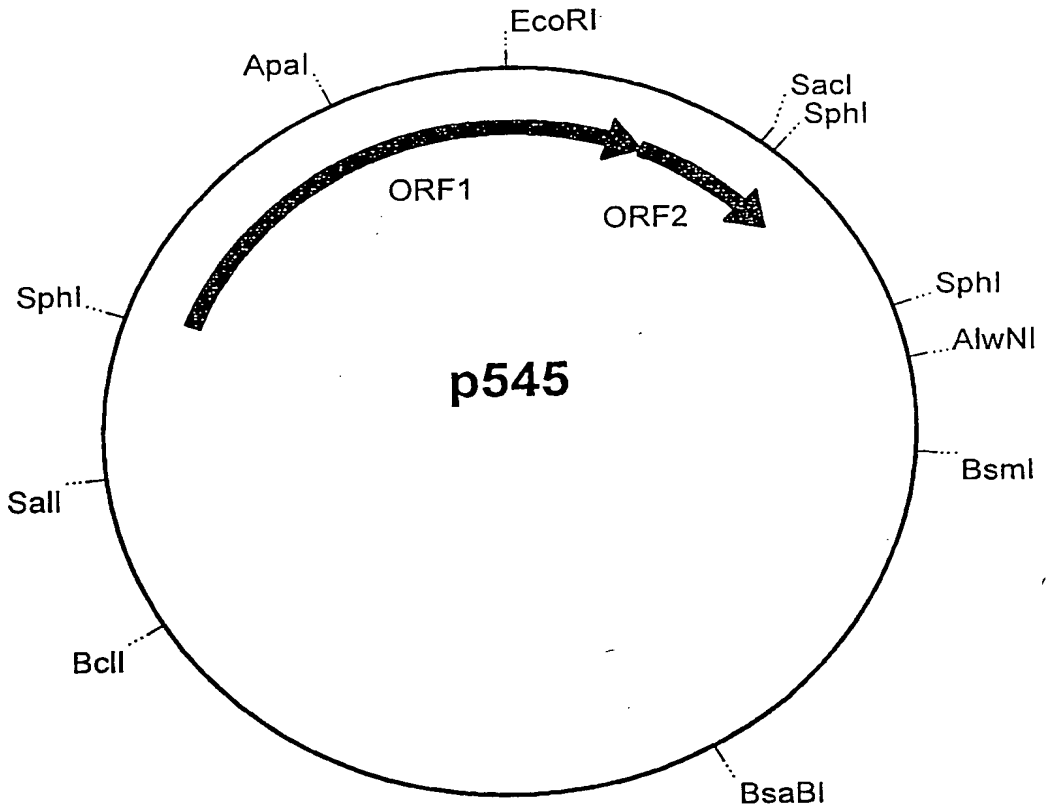
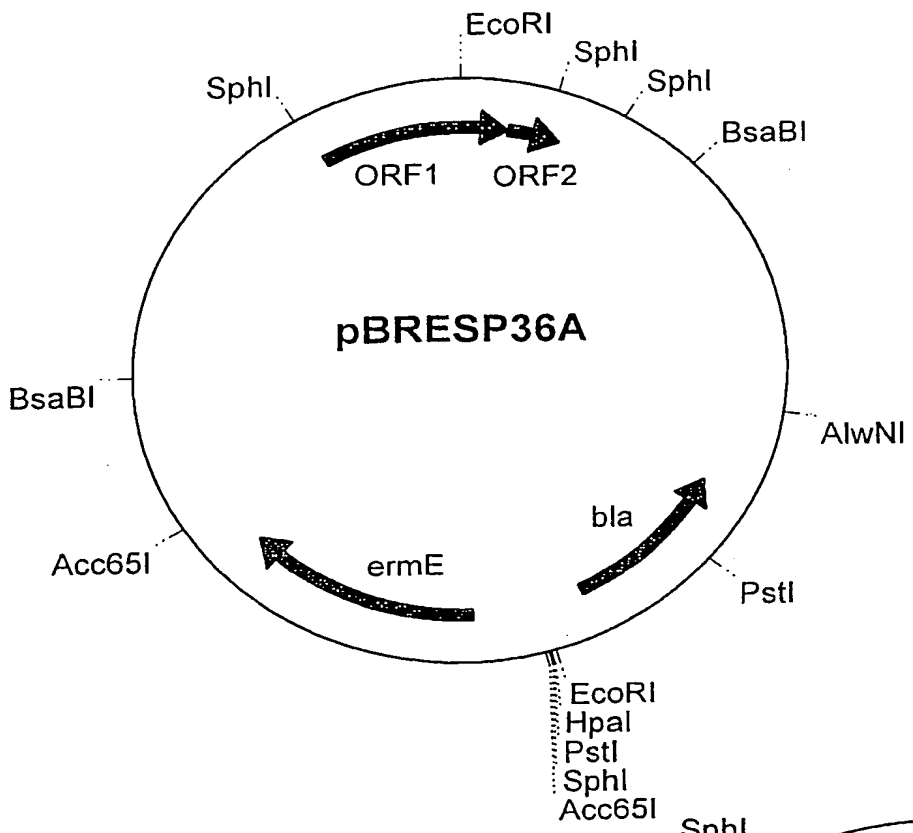


Figure 2a



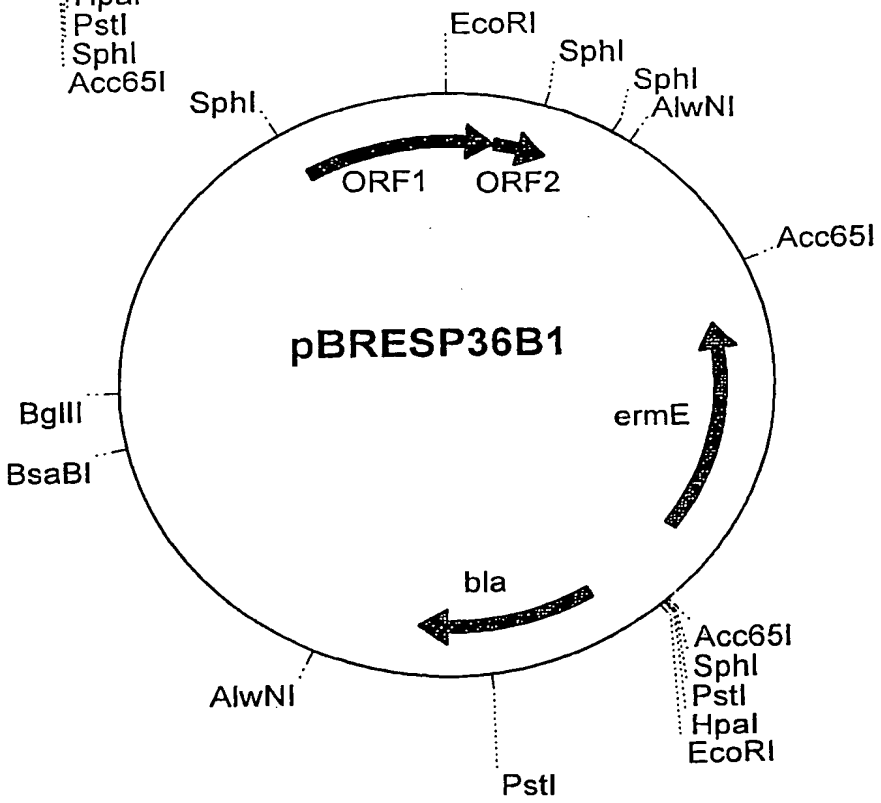
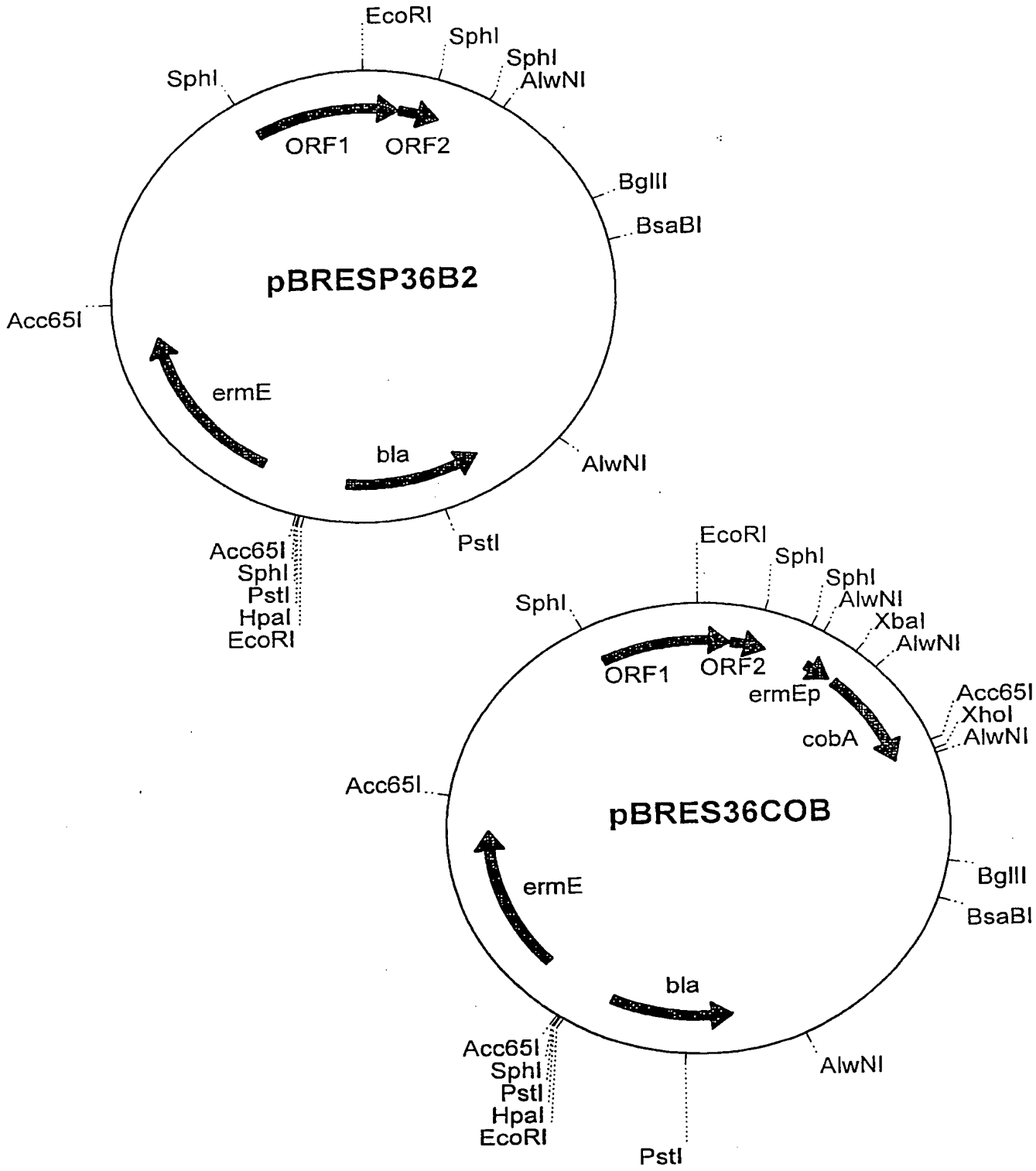
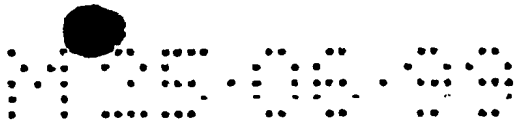


Figure 2b





```
2875  atggattcgt tcgagacgtt gttccctgag agctggctgc cacgcaagcc gctggcgtca
        M  D  S   F  E  T   L  F  P  E   S  W  L   P  R  K   P  L  A  S

2935  gccgagaagt ctggggcgta ccggcacgtg actcggcaga gggcgctgga gctgccttac
        A  E  K   S  G  A   Y  R  H  V   T  R  Q   R  A  L   E  L  P  Y

2995  atcgaagcga acccgttggt catgcagtcc ttggtcatca ccgatcgaga tgcttcggat
        I  E  A   N  P  L   V  M  Q  S   L  V  I   T  D  R   D  A  S  D

3055  gctgactggg ccgcagacct cgctgggctg ccttcaccgt cctacgtgtc catgaaccgt
        A  D  W   A  A  D   L  A  G  L   P  S  P   S  Y  V   S  M  N  R

3115  gtcacgacca ccggacacat cgtctatgcc ttgaagaacc ctgtgtgtct gaccgatgcc
        V  T  T   T  G  H   I  V  Y  A   L  K  N   P  V  C   L  T  D  A

3175  gcgcggcgac ggcctatcaa cctgctcgcc cgcgtcgagc agggcctatg cgacgttctc
        A  R  R   R  P  I   N  L  L  A   R  V  E   Q  G  L   C  D  V  L

3235  ggcggcgatg catcctacgg gcaccggatc acaaagaacc cgctcagcac cgcccatgcg
        G  G  D   A  S  Y   G  H  R  I   T  K  N   P  L  S   T  A  H  A

3295  accctctggg gccccgcaga cgcgctctac gagctgcgcg ccctcgcaca caccctcgac
        T  L  W   G  P  A   D  A  L  Y   E  L  R   A  L  A   H  T  L  D

3355  gagatccacg cactgccgga ggcagggaac ccgcgtcgca acgtcacccg atcaacggtc
        E  I  H   A  L  P   E  A  G  N   P  R  R   N  V  T   R  S  T  V

3415  ggccgcaacg tcaccctgtt cgacaccacc cgcatgtggg cataccgggc cgtccggcac
        G  R  N   V  T  L   F  D  T  T   R  M  W   A  Y  R   A  V  R  H

3475  tcctggggcg gcccggtcgc cgaatgggag cacaccgtat tcgagcacat ccacctactg
        S  W  G   G  P  V   A  E  W  E   H  T  V   F  E  H   I  H  L  L

3535  aacgagacga tcatcgccga cgaattcgcc acaggccccc tcggcttgaa cgaacttaag
        N  E  T   I  I  A   D  E  F  A   T  G  P   L  G  L   N  E  L  K

   0  cacttatctc gatccatttc ccgatgggtc tggcgcaact tcacccccga aaccttccgc
        H  L  S   R  S  I   S  R  W  V   W  R  N   F  T  P   E  T  F  R

 100  gcacgccaga aagcgatcag cctccgtgga gcatccaaag gcggcaaaga aggcggccac
        A  R  Q   K  A  I   S  L  R  G   A  S  K   G  G  K   E  G  G  H

 160  aaaggcggca ttgccagtgg cgcatcacgg cgcgcccata cccgtcaaca gttcttggag
        K  G  G   I  A  S   G  A  S  R   R  A  H   T  R  Q   Q  F  L  E

 220  ggtctctcat ga
        G  L  S   -
```

Figure 3

```
228  atgaccacac gtgaacgtct cccccgcaac ggctacagca tcgccgctgc tgcgaaaaag
       M  T  T   R  E  R   L  P  R  N   G  Y  S   I  A  A   A  A  K  K

288  ctcggtgtct ccgagtccac cgtcaagcgg tggacttccg agccacgcga ggagttcgtg
       L  G  V   S  E  S   T  V  K  R   W  T  S   E  P  R   E  E  F  V

348  gcccgcgttg ccgcacgcca cgcgcggatt cgtgagctcc gctcggaggg tcagagcatg
       A  R  V   A  A  R   H  A  R  I   R  E  L   R  S  E   G  Q  S  M

408  cgtgcgattg ctgccgaggt cggggtttcc gtgggcaccg tgcactacgc gctgaacaag
       R  A  I   A  A  E   V  G  V  S   V  G  T   V  H  Y   A  L  N  K

468  aatcgaactg acgcatga
       N  R  T   D  A  -
```

Figure 4

ABSTRACT

PROPIONIBACTERIUM VECTOR

An endogenous plasmid of Propionibacterium is described, isolated from Propionibacteria freudenreichii LMG 16545 (deposited as CBS 101022), and its sequence provided. This plasmid can be used to transform Propionibacteria to express homologous or heterologous proteins, in the production of recombinant proteins or products of enzymes, for example vitamin B12.